ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Discrete Applied Mathematics 128 (2003) 75–83

DISCRETE APPLIED MATHEMATICS

www.elsevier.com/locate/dam

# Intersecting codes and separating codes

G. Cohnen[a], S. Encheva[b,*], S. Litsyn[c], H.G. Schaathun[d,1]

[a] *Ecole Nationale Supérieure des Télécommunications, 46 rue Barrault, 75634 Paris, France*
[b] *Stord/Haugesund College, Bjørnsonsg. 45, 5528 Haugesund, Norway*
[d] *EES Dept., Tel Aviv University, 69978 Ramat Aviv, Israel*
[c] *Dept. of Informatics, UiB, HIB, N-5020, Bergen, Norway*

Received 19 February 2001; received in revised form 20 December 2001; accepted 8 April 2002

**Abstract**

Let $\Gamma$ be a code of length $n$. Then $x$ is called a descendant of the coalition of codewords $a, b, \ldots, e$ if $x_i \in \{a_i, b_i, \ldots, e_i\}$ for $i = 1, \ldots, n$. We study codes with the following property: any two non-intersecting coalitions of a limited size have no common descendant.

We present constructions based on linear intersecting codes.



*Keywords:* Intersecting code; Separating code; Copyright protection

## 1. Introduction

Let us start by mentioning two new problems which were a motivation for studying separating codes.

Consider the distribution of digital content to subscribers. Each authorized user is given a decoder (e.g. a smartcard) with a secret decryption key. The distributor broadcasts an encrypted version of the content, which is decrypted by the authorized users. The scope of applications encompasses watermarking and fingerprinting issues, as well as pay-per-view television, e-commerce and any broadcasting system to subscribers.

Another application is Digital Fingerprinting: suppose a Distributor wishes to create and distribute a large number of copies of a file. In order to trace illegal copies he will

* Corresponding author. Tel.:+47-52702685; fax: +47-52702601.

*E-mail addresses:* cohen@inf.enst.fr (G. Cohnen), sbe@hsh.no (S. Encheva), litsyn@eng.tau.ac.il (S. Litsyn), georg@ii.uib.no (H.G. Schaathun).

[1] Part of the work was done at the Ecole Nationale Supérieure des Télécommunications in Paris. Schaathun had his stay supported by The Norwegian Research Council under Grant 138654/410.

0166-218X/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
PII: S0166-218X(02)00437-7

mark each one, by changing a few elements of the file belonging to some subset of a privileged set of coordinates called marks. The subset of marks associated to a copy is called a fingerprint. A collusion occurs when a coalition of $t$ pirate users compare their fingerprinted copies: whenever they differ on some coordinate they will know it is a mark. They can then produce an illegal copy by changing elements on the subset of marks they have found out. Following previous work, we suppose that they cannot access the other marks.

In both instances, codes were studied (see [5,3]) as a method to prevent a coalition of a given size from forging some type of copy. Among the forbidden moves, let us mention: framing another user (frameproof codes), getting away with no member of the coalition being caught (identifying codes, studied for coalitions of size 2 in [9] and in [2]) for larger coalitions.

A first step in identification is to forbid disjoint coalitions from producing the same copy or decoder. This turns out to have been studied in another context under the name of "separation" (see [15,14] for a long Saga of pioneering contributions); see also [8,10].

In this paper, we present bounds and efficient constructions for separating codes based on linear intersecting codes.

## 2. Definitions

For any positive real number $x$ we denote by $\lceil x \rceil$ the smallest integer at least equal to $x$, and by $\lfloor x \rfloor$ the largest integer at most equal to $x$. A subset $\Gamma$ of $GF(q)^n$, the vector space of dimension $n$ over the finite field with $q$ elements $GF(q)$, is called an $(n, M, d)$-*code* if $|\Gamma| = M$ and the minimum Hamming distance between two of its elements (codewords) is $d$.

Consider $\mathfrak{I} \subseteq \Gamma$. For any position $i$ define the *projection* $P_i(\mathfrak{I}) = \bigcup_{a \in \mathfrak{I}} a_i$. Define the *feasible set* of $\mathfrak{I}$ by

$$F(\mathfrak{I}) = \{x \in GF(q)^n : \forall i, x_i \in P_i(\mathfrak{I})\}.$$

The feasible set $F(\mathfrak{I})$ represents the set of all possible $n$-tuples (descendants) that could be produced by the coalition $\mathfrak{I}$ by comparing the codewords they jointly hold. Observe that $\mathfrak{I} \subseteq F(\mathfrak{I})$ for all $\mathfrak{I}$.

If two non-intersecting coalitions can produce the same descendant, it will be impossible to trace with certainty even one pirate. This motivates the following reworded definition from [8].

**Definition 1.** A code $C$ is $(t, t')$-separating if, for any pair $(T, T')$ of disjoint subsets of $C$ where $|T| = t$ and $|T'| = t'$, the feasible sets are disjoint, i.e. $F(T) \cap F(T') = \emptyset$.

Since the identification property is preserved by translation, we shall always assume that $\mathbf{0} \in \Gamma$.

The identification property can be rephrased as follows when $q=2$: for any ordered $2t$-tuple of codewords, there is a coordinate where the $2t$-tuple (1..10..0) of weight $t$ or its complement occurs.

We denote by $C[n,k,d]_q$ (or simply $C[n,k]_q$ when $d$ is irrelevant) a *linear* code (i.e., a vectorial subspace) of length $n$, dimension $k$ over $GF(q)$ and minimum distance $d$. The *rate* of $C$ is $R(C)=R=k/n$. In the non-linear case, the rate is defined analogously as $n^{-1}\log_q M$. We refer to [11] for all undefined notions on codes.

## 3. Intersecting codes

**Definition 2.** A linear code of dimension $k \geqslant t$ is said to be $t$-wise intersecting if any $t$ linearly independent codewords have intersecting supports.

For results and constructions of intersecting codes, see, e.g., [7]. Connections between intersecting codes have implicitely been made for the cases $t=2,3$. We summarize them in the next result.

**Proposition 1.** *For a binary linear code, the following properties are equivalent*:

(1) $(2,1)$-*separation and* 2-*wise intersection* [12];
(2) $(2,2)$-*separation and* 3-*wise intersection* [4].

The goal of this section is to consider higher values of $t$. First we give a partial extension of the previous result to the $q$-ary case:

**Proposition 2.** *Every linear* $(2,2)$-*separating* $[n,k]$*code with* $k \geqslant 3$ *is* 3-*wise intersecting*.

**Proof.** If $k \leqslant 2$, the proposition holds trivially, so assume that $k \geqslant 3$. Suppose $C$ is $(2,2)$-separating, and consider three independent codewords $a,b,c$. We shall prove that these three words have intersecting supports. Consider the $(2,2)$-configuration $(\mathbf{0}, c+a; a, b)$. Since $C$ is $(2,2)$-separating, there is a position $i$ where $a$ is $\alpha \neq 0$ and $b$ is $\beta \neq 0$, and $c+a$ is $\gamma \notin \{\alpha,\beta\}$. Now $c$ is $\gamma-\alpha \neq 0$ on position $i$. □

**Example 1.** The 3-wise binary intersecting $[126,14]$ code [7], yields a $(2,2)$-linear separating code with parameters $(126,2^{14})$ (already in [15]).

The asymptotical (in $n$) existence of 3-wise intersecting codes with rate $1-(1/3)\log_2 7$ is shown in [7]. This gives a *linear* (2,2)-separating code with a rate already achieved in [15] by different methods.

**Proposition 3.** *If* $C$ *is a* $t$*-wise intersecting binary linear code, and* $\Gamma \subseteq C$ *is a subset such that any* $t$ *of its elements are linearly independent, then* $\Gamma$ *is* $(j, t+1-j)$*-separating for all* $j$ *such that* $1 \leqslant j \leqslant t$.

**Proof.** Choose any (two-part) sequence $Y'$ of $t+1$ codewords from $\Gamma$,

$$Y' := (a'_1, \ldots, a'_j; c'_1, \ldots, c'_{t+1-j}).$$

$Y'$ is $(j, t+1-j)$-separated if and only if $Y := Y' - c'_{t+1-j}$ is. Hence it suffices to show that

$$Y = (a_1, \ldots, a_j; c_1, \ldots, c_{t-j}, \mathbf{0})$$

is $(j, t+1-j)$-separated.

Since any $t$ codewords in $Y'$ are linearly independent, so are the $t$ first codewords of $Y$.

Now, consider

$$\{a_1 + c_1, \ldots, a_1 + c_{t-j}; a_1, \ldots, a_j\},$$

which is, by linear algebra, a set of linearly independent codewords from $C$, and hence all equal to 1 on some coordinate $i$. Since $a_1 + c_l$ is 1 on coordinate $i$, $c_l$ must be zero for all $l$. Hence $Y$, and consequently $Y'$, is separated on coordinate $i$. □

**Proposition 4.** *If $C$ is a $t$-wise intersecting binary linear code, and $\Gamma \subseteq C$ is such that any $t-1$ of its elements are linearly independent, then $\Gamma$ is $(j, t+1-j)$-separating for all even $j$ such that $1 < j \leqslant t$.*

**Proof.** We define $Y$ as in the previous proof, and the $t-1$ first codewords of $Y$ are linearly independent. If $c_{t-j}$ is linearly independent of the others, then we are done by the first proof; hence we assume that $c_{t-j}$ is dependent on the $t-1$ first codewords, and since any $t-1$ codewords are independent, it must in fact be the sum of the $t-1$ first codewords. By the same argument as in the previous proof, we get one coordinate $i$, where $a_1 + c_1, \ldots, a_1 + c_{t-1-j}, a_1, \ldots, a_j$ are all one, and $c_1, \ldots, c_{t-1-j}$ are zero. Now, $c_{t-j}$ is the sum of the $t-1$ first codewords, of which $j$ are 1 and the rest are zero on coordinate $i$. Since $j$ is even, $c_{t-j}$ is zero, and $Y$ is separated. □

Note that if $t$ is even, then either $j$ or $t+1-j$ is even; thus we get the following corollary.

**Corollary 1.** *If $C$ is a binary linear $t$-wise intersecting code, $t$ is even and $\Gamma \subseteq C$ is a subset such that any $t-1$ of its elements are linearly independent, then $\Gamma$ is $(j, t+1-j)$-separating for all $j$ such that $1 \leqslant j \leqslant t$.*

The rest of the section is devoted to proving that, given a $t$-wise intersecting code, a nonlinear subcode with the prescribed properties and a certain rate does in fact exist.

**Lemma 1.** *Given an $[n, rm+1]$ linear, binary code $C$, we can extract a non-linear subcode $\Gamma$ of size $2^r$ such that any $2m+1$ codewords are linearly independent.*

Note that the rate of $\Gamma$ is approximately $R/m$ where $R = (rm+1)/n$ is the rate of $C$.

**Proof.** Let $C'$ be the $[2^r, 2^r - 1 - rm, 2m + 2]$ extended BCH code. The columns of the parity check matrix of $C'$ make a set $\Gamma'$ of $2^r$ vectors from $GF(2)^{rm+1}$, such that any $2m + 1$ of them are linearly independent. Now there is an isomorphism $\phi : GF(2)^{rm+1} \to C$, so let $\Gamma = \phi(\Gamma')$. □

**Theorem 1.** *Given an* $[n, nR]$ *t-wise intersecting binary code with* $t \geqslant 3$, *there is a construction of a non-linear code* $\Gamma$ *of rate approximately* $R/\lfloor (t-1)/2 \rfloor$, *which is* $(j, t+1-j)$*-separating.*

**Proof.** First consider $t$ even, and write $t = 2m + 2$, where $m \geqslant 1$. By Corollary 1, we want to extract $\Gamma$ such that any $2m + 1$ codewords are independent, and such $\Gamma$ exists with rate $R/m$ by Lemma 1.

Then consider odd $t$, and write $t = 2m + 1$, where $m \geqslant 1$. By Proposition 3, we want to extract $\Gamma$ such that any $2m + 1$ codewords are independent, and such $\Gamma$ exists with rate $R/m$ by Lemma 1. □

**Example 2.** In [7], it was shown that for sufficiently large $n$, and for any rate $R < 1 - (1/t)\log(2^t - 1)$, there are $t$-wise intersecting linear, binary $[n, k]$ codes of rate $R$. Though non-constructive, this result guarantees the existence, for any $t \geqslant 3$, of non-linear, binary codes which are $(j, t+1-j)$-separating for all $j$ and have rates arbitrarily close to

$$\frac{1 - (1/t)\log(2^t - 1)}{\lfloor (t-1)/2 \rfloor}.$$

Note that random methods (see [1]) give a better rate of $1 - (1/t)\log(2^t - 1)$. Our method, though, can be made constructive if constructions of intersecting codes are used.

## 4. Constructions

We will give some construction in the binary and ternary cases. In addition to the results from the previous section, we need a couple of preliminaries from previous papers.

The following classical coding method (known as *concatenation*, see e.g. [11]) is quite powerful to obtain $p$-ary separating codes from $q$-ary ones, $q = p^k$. We state it in the linear version, although it can easily be rephrased in the nonlinear case.

Let $C_1$ be an $[N, K, D]_q$ code over $GF(q)$, $q = p^k$; let $C_2$ be an $[n, k, d]_p$ $p$-ary code. We map (by an isomorphism of additive groups) $GF(q)$ onto $GF(p)^k$, and then associate to $\alpha \in GF(2^k)$ the codeword $c(\alpha) = \alpha G$ of $C_2$, where $G$ is a generator matrix of $C_2$.

Denoting by $C_1 \bigstar C_2$ the concatenation of $C_1$ and $C_2$, we have the following easy result (see [15]):

**Proposition 5.** *$C_1 \star C_2$ is an $[Nn, Kk, Dd]$ p-ary code. If $C_1$ and $C_2$ are both $(t,t)$-separating codes (over $GF(q)$ and $GF(p)$, respectively), then $C_1 \star C_2$ is a $(t,t)$-separating p-ary code.*

Concatenation is useful when combined with the next result, which provides a sufficient condition for a code to be separating, solely based on its minimum distance.

**Proposition 6.** *Let $\Gamma$ be a code with $d/n > 1 - 1/t^2$; then it is a $(t,t)$-separating code.*

In fact, the condition $d/n > 1 - 1/t^2$ guarantees a much stronger property: *t-traceability* [5,16], namely that all closest codewords to the produced descendant are part of the coalition producing it. It thus insures the identifiable parent property of [9], with the extra feature of a search algorithm linear in $|\Gamma|$.

For $t = 2$, the weaker condition $4d > 3D$ is enough for a linear code to be (2,2)-separating, where $D$ denotes the largest code distance (see Chap. 7 of [14,15] for the binary case, and [6] for the general case).

### 4.1. Binary constructions

We now combine concatenation with the following result to construct infinite families of separating binary codes. This was done by Sagalovitch for $(2,1)$ and $(2,2)$ separation.

**Theorem 2** (Tsfasmann [17]). *For any $\alpha > 0$ there is an infinite families of codes $\mathcal{U}(N)$ with parameters $[N, NR, N\delta]_q$ for $N \geqslant N_0(\alpha)$ and*

$$R + \delta \geqslant 1 - (\sqrt{q} - 1)^{-1} - \alpha.$$

**Proposition 7** (Cohen and Zémor [7]). *The punctured dual of the 2-error-correcting BCH code with parameters $[2^{2t+1} - 2, 4t + 2, 2^{2t} - 2^t - 1]_2$ is t-wise intersecting.*

**Example 3.** For $t = 4$, we get from Proposition 7 a 4-wise intersecting code with parameters $[2^9 - 2, 18]_2$. Now the subset $\Gamma'$ of the $2^{17}$ codewords having a 1 in the last position (say) is clearly such that any 3 of its elements are independent, thus we get a $(3,2)$-separating $(2^9 - 3, 2^{17})$ code by Corollary 1. We can concatenate $\Gamma$ with the code $\mathcal{U}(N)$ with parameters $[N, RN, 5N/6+1]_{2^{18}}$ from Theorem 2 to get $(3,2)$-separating codes with rates $R \approx 0.00557$.

The previous example provides a method for shortening:

If $\Gamma(n, M)$ is $(t,t')$-separating, then so are the 2 subcodes $\Gamma_0$ (resp. $\Gamma_1$) having 0 (resp. 1) in the first coordinate. Taking the largest and removing the first coordinate (which no longer separates anything), gives a shortened $(n-1, \lceil M/2 \rceil)$ $(t,t')$-separating code.

**Proposition 8.** *There is a constructive infinite sequence of binary* $(j, t+1-j)$*-separating codes of rate* $2^{-3(t-1)}(1+\mathrm{o}(1))$.

This proposition follows directly from the following lemma:

**Lemma 2** (Cohen and Zémor [7]). *There is a constructive infinite sequence of* $t$*-wise intersecting binary codes with rate arbitrarily close to*

$$R_t = \left(2^{1-t} - \frac{1}{2^{2t+1}-1}\right)\frac{2t+1}{2^{2t}-1} = 2^{2-3t}(t+\mathrm{o}(t)).$$

**Proof.** By concatenating geometric $[N,K,D]_q$ codes from Theorem 2 satisfying $D > N(1-2^{1-t})$ with $q = 2^{4t+2}$ and rate arbitrarily close to $2^{1-t} - 1/(\sqrt{q}-1)$, with the $[2^{2t+1}-2, 4t+2, 2^{2t}-2^t-1]$ code of Proposition 7, we obtain the result. □

**Example 4.** Let $q = p^{2m}$. Consider (see Theorem 2) a family of codes $\mathfrak{V}(N)$ with parameters $[N, NR, N\delta]_q$ with $N \geqslant N_0(\alpha)$ and

$$R + \delta \geqslant 1 - (p^m - 1)^{-1} - \alpha.$$

Choosing $p = 2, m = 7$, $\delta = 3/4 + \varepsilon$, (see Proposition 6) and concatenating $\mathfrak{V}(N)$ and $C$, the binary $[126, 14, 55]$ code, yields a constructive infinite sequence $\{\mathfrak{V}(N) \circ C\}_N$ of binary linear $(2,2)$-separating codes with rates arbitrarily close to 0.026.

### 4.2. Ternary constructions

The ternary construction will make use of three codes, and apply twice the concatenation method.

The first seed $C_1$ is the $[4,2,3]_3$ tetracode (see for example [13]). This code is self-dual, MDS (on Singleton's bound $d = n - k + 1$). It is both an extended perfect Hamming code and a simplex (all codewords are at distance 3 apart). A basis of the $[4,2,3]_3$ code is $\{1110, 0121\}$. It is (2,2)-separating (in fact, it is even 2-traceable, see [9]).

The second seed we use to concatenate with the tetracode is the extended Reed-Solomon code $C_2[9,3,7]_{3^2}$. It is $(2,2)$-separating by Proposition 6. The result is $C_1 \bigstar C_2[36,6]_3$ which is a $(2,2)$-separating by Proposition 5.

Now this code is a large enough seed for the algebraic-geometry codes of [17] (see Theorem 2) to work efficiently.

By concatenation with an $[N, K, D = \lceil 3N/4 \rceil + 1]_{3^6}$ algebraic-geometry code $C(N)$ of rate approximately $\frac{1}{4} - (3^3-1)^{-1}$, this gives a constructive family $\{C_1 \bigstar C_2 \bigstar C(N)\}_N$ of linear ternary (2,2)-separating codes with rate $R \approx \frac{11}{312}$.

## 5. Upper bounds on intersecting codes

We now present upper bounds on the rate of such codes, based on projection arguments analoguous to those of [15].

**Theorem 3.** *A $t$-wise intersecting code $C_t[n,k,d]$ gives rise by projection to a $(t-1)$-wise intersecting code $C_{t-1}[d,k-1]$.*

**Proof.** Let $a \in C$ be a fixed element of minimum weight $d$. Denote by $C_a$ the $[n,k-1]$ supplementary subspace of $\{\mathbf{0},a\}$ in $C$. Consider any $(t-1)$ independent codewords $\{b^1,\ldots,b^{t-1}\}$ in $C_a$. Then $\{a,b^1,\ldots,b^{t-1}\}$ is full rank, hence these $t$ codewords of $C$ intersect (on the support of $a$). Thus $C/a$, the projection of $C_a$ on the support of $a$ is a $(t-1)$-intersecting $[d,k-1]$ code. □

To get an upper bound on the dimension of such codes in the binary case, we use recursively any upper bound from coding theory, for instance the McEliece et al. bound (see [11]):

$$R \leqslant H_2\left(\frac{1}{2} - \sqrt{\frac{d}{n}\left(1-\frac{d}{n}\right)}\right).$$

For $t=3$, we get the following sequence of codes:

$$C_3[n,k,d], \quad C_2[d,k-1,d'], \quad C_1[d',k-2],$$

where $C_i$ is $i$-wise intersecting, and has rate $R_i$.

Considering $C_1$, we have that $k-2 \leqslant d'$, which implies that

$$R_2 = (k-1)/d \leqslant (d'-1)/d \leqslant d'/d.$$

By the McEliece bound, this implies $R_2 \leqslant 0.28$. Finally we have

$$R_1 = \frac{k}{n} \leqslant \frac{0.28d+1}{n} \leqslant 0.108,$$

where the final bound follows by applying again the McEliece bound. Note that the same bound holds for linear $(2,2)$-separating codes (see [15]), and these codes are equivalent to 3-wise intersecting codes by Theorem 1.

The following corollary arises from the same technique and some other values for $t$.

**Corollary 2.** *The asymptotic rate of the largest $t$-wise intersecting code is at most $R_t$, with $R_2 \approx 0.28, R_3 \approx 0.108, R_4 \approx 0.046, R_5 \approx 0.021, R_6 \approx 0.0099$.*

## Acknowledgements

We thank the referees for careful reading and comments and Grisha Kabatiansky for numerous friendly constructive discussions.

## References

[1] A. Barg, G.R. Blakeley, G. Kabatiansky, Good digital fingerprinting codes, Proceedings IEEE ISIT, Washington, DC, 2001, p. 161.

[2] A. Barg, G. Cohen, S. Encheva, G. Kabatiansky, G. Zémor, A hypergraph approach to the identifying parent property, SIAM J. Discrete Math. 14 (2001) 423.
[3] D. Boneh, J. Shaw, Collusion-secure fingerprinting for digital data, Springer Lecture Notes in Computer Science, Vol. 963, Springer, Berlin, 1995, p. 452.
[4] B. Bose, T.R.N. Rao, Separating and completely separating systems and linear codes, IEEE Trans. Comput. 29 (1980) 665.
[5] B. Chor, A. Fiat, M. Naor, Tracing traitors, Springer Lecture Notes in Computer Science, Vol. 839, Springer, Berlin, 1994, p. 257.
[6] G. Cohen, S. Encheva, H.-G. Schaathun, More on $(2,2)$-separating systems, IEEE Trans. Inform. Theory, in print.
[7] G. Cohen, G. Zémor, Intersecting codes and independent families, IEEE Trans. Inform. Theory 40 (1994) 1872.
[8] A.D. Friedman, R.L. Graham, J.D. Ullman, Universal single transition time asynchronous state assignments, IEEE Trans. Comput. 18 (1969) 541.
[9] H.D.L. Hollmann, J.H. van Lint, J.-P. Linnartz, L.M.G.M. Tolhuizen, On codes with the identifiable parent property, J. Combin. Theory Ser. A 82 (1998) 121.
[10] J. Körner, G. Simonyi, Separating partition systems and locally different sequences, SIAM J. Discrete Math. 1 (1988) 355.
[11] F.J. MacWilliams, N.J.A. Sloane, The Theory of Error-Correcting Codes, North-Holland, Amsterdam, 1977.
[12] D.R. Pradhan, S.M. Peddy, Techniques to construct $(2,1)$ separating systems from linear error-correcting codes, IEEE Trans. Comput. 25 (1976) 945.
[13] E.M. Rains, N.J.A. Sloane, Self-dual codes, in: Handbook of Coding Theory, North-Holland, Amsterdam, 1998, p. 177.
[14] Yu.L. Sagalovich, State Encoding and Reliability of Automata, Svyaz', Moscow, 1975 (in Russian).
[15] Yu.L. Sagalovich, Separating systems, Problems Inform. Transmission 30 (1994) 105.
[16] D.R. Stinson, R. Wei, Combinatorial properties and constructions of traceability schemes and frameproof codes, SIAM J. Discrete Math. 11 (1998) 41.
[17] M.A. Tsfasmann, Algebraic–geometric codes and asymptotic problems, Discrete Appl. Math. 33 (1991) 241.